A METHOD AND SCHEDULER FOR PERFORMING A SCHEDULING ALGORITHM WITH MINIMUM RESOURCE PARAMETER

This invention relates to communication systems. It is particularly applicable to wireless communication system where a scheduling algorithm assigns channel resources to users for transmission and reception of data. More specifically, the present invention relates to a method and scheduler for performing a scheduling algorithm as set forth in the independent claims.

In a cellular mobile communication system, the mobile station usually transmits and receives information to and from a base station using channel resources such as time slots, frequency bandwidth, code sequence, or combinations thereof. These resources are generally shared among users of the communication system.

In an ad-hoc mobile communication system, a radio access point usually transmits and receives information to and from other radio access points within the same ad-hoc network, using channel resources such as time slots, frequency bandwidth, code sequence, or combinations thereof. These resources are generally shared among users of the communication system. In such an ad-hoc network, there may either be a dedicated master access point for managing the ad-hoc network, or alternatively one radio access point may adopt the master access point functionalities for management of the ad-hoc network.

It should be apparent to those skilled in the art that a cellular base station and an ad-hoc master access point share at least part of the responsibilities for managing the resources and users within their area of coverage. Similarly the role of a radio access point in an ad-hoc network has a lot in common to the role of a mobile device in a cellular wireless system. For matters of simplicity the following description shall be referring to a cellular wireless network. Changes required for application of the present invention to ad-hoc networks are easily derived for those skilled in the art from the description.

In the context of wireless communication, all operations involved in either transmitting or receiving data are referred to as processing data. In order to process data, the

mobile device has to expend operating and processing power for its equipment. From an economical point of view, the most reasonable expenditure of this power is obtained if in return the mobile receiver processes a lot of data while the power is spent. On the other hand, such power is rather wasted if while expending the power no or little data is processed.

The allocation of data to a user via the channel resource is usually done by a scheduling algorithm. At least for downlink, i.e. for the transmission direction from base station to mobile terminal, such a scheduler is usually operating in the base station or other parts of the non-mobile entities within the communication system. Such a scheduler usually evaluates parameters such as service data rate, channel state, but does not take into account an economic factor as described above. Even for uplink, i.e. for the transmission direction from mobile terminal to base station, a scheduler may operate in a central node (e.g. base station in cellular systems, master station in an ad-hoc network) to allocate resources. The result of such central node scheduling may then be transmitted to the mobile entities.

In wireless communication systems employing Dynamic Channel Assignment (DCA) schemes air-interface resources are assigned dynamically to links between a base station (BS) and multiple mobile terminals (MT). A layout of a typical communication system is shown in Fig. 1, wherein a BS serves several MTs in a service area. The air-interface resources are usually defined by a logical channel, where a logical channel corresponds to e.g. one or multiple codes in a CDMA system, one or multiple subcarriers in an OFDM system, one or multiple timeslots in a TDMA system (e.g. GSM), or to combinations of those e.g. in an OCDMA or an MC-CDMA system. DCA can be applied to uplink and downlink.

Employing Adaptive Modulation and Coding (AMC), the data-rate within a scheduling frame for a scheduled MT will be adapted to the instantaneous channel quality of the respective link by changing the modulation and coding scheme dynamically. AMC is typically applied jointly with DCA.

In a system making use of DCA and AMC a so-called scheduler decides which resources are assigned to which MT. A commonly used approach is to use centralized scheduling, where the scheduler is located in the BS and performs its

decision based on the following side information, such as channel quality information of the links to the MTs or offered traffic for specific links e.g. amount of data available for transmission to a specific MT.

Common objectives of the scheduler are to achieve fairness between users, maximize system throughput and/or fulfill Quality of Service (QoS) requirements (e.g. delay, data-rate, loss rate, jitter) for the services run by the scheduled mobile terminals. In state-of-the-art wireless communication systems the scheduler works on a packet basis.

The following schedulers are well known examples in the area of wireless communications:

- Round Robin (RR) Scheduler:

  This scheduler allocates equal air-interface resources to all MS independent of the channel conditions thus achieving fair sharing of resources.

- Max-Rate (MR) or Max C/I (MC) Scheduler:

  This scheduler chooses the user with the highest possible instantaneous data-rate (carrier-to-interference C/I ratio). It achieves the maximum system throughout but ignores the fairness between users.

- Proportional Fair (PF) Scheduler :

  This scheduler maintains an average data-rate transmitted to each user within a defined time window and examines the ratio of the instantaneous to the average channel conditions (or ratio of the instantaneous possible data-rate to the average data-rate) experienced by different users and chooses the user with the maximum ratio. This scheduler increases the system throughput with respect to RR scheduling, while maintaining some degree of long-term fairness.

More detailed information on the structure and function of a scheduler can be obtained for example from US 2003/0104817 which discloses a method for scheduling multiple users sharing a communication resource, particularly relating to high data rate wireless transmission putting emphasis on QoS considerations.

In current systems, a terminal may transmit signals to tell the scheduler what data rate is necessary to satisfy the user or service. Among other parameters, this may involve an average connection (or service) data rate and a maximum allowable delay. However, the scheduling at BS cannot know whether an MT is working efficiently in terms of power consumption for reception of data.

The object underlying the present invention is to provide a method and scheduler for performing a scheduling algorithm with minimum resource scheduling.

This object is solved by a method and a scheduler as set forth in the independent claims.

The invention is based on the recognition that from the viewpoint of the MT in an effort to work economically reasonable, it is either preferred that allocation units are scheduled to a user or a service such that either a minimum resource constraint is satisfied or that the allocation units scheduled for that user for a particular scheduling frame are released, i.e. no data is transmitted to that specific user in order to enable the MT, for instance to switch off receiver circuitry in order to save power resources.

A scheduler has information about QoS parameters and other constraints that have influence of the assignment of allocation units (AU) to an MT. Among those parameters, there exists a long-term average or minimum resource figure to meet the overall satisfaction demands, as well as a short-term minimum resource figure that represents a number of resources that should at least be assigned to the user in a scheduling frame. In case this short-term figure is not met by the scheduling algorithm, the preferred solution is to free the AUs for that user in that particular scheduling frame and to wait for subsequent scheduling frames when more resources can be assigned to the user.

According to a preferred embodiment, the scheduling step includes considering at least one or a combination of the scheduling parameters channel condition, amount of data available for transmission to a specific user, quality of service, delay, data rate or a carrier to interference ratio.

Preferably, the scheduling frame has at least one or a combination of a time division, frequency or code division frame structure.

According to one embodiment, the resource constraint is a long-term user or service based requirement which may be monitored for several scheduling frames.

According to a further embodiment, the resource constraint is a short-term frame-based requirement and consequently monitors instantaneous resource constraints.

Preferably, the allocation units scheduled for a user or service have a quantity of either transmittable information bits, internet protocol packets, code blocks or modulation symbols.

According to a preferred embodiment, the scheduling algorithm comprises the additional step of checking whether at least one other resource constraint is not violated by releasing allocation units. Only if a violation by such release does not occur, the allocation units are released. In this way, it is ensured that other more urgent long-term based requirements are fulfilled even if the short-term frame-based requirement is not met.

According to an advantageous embodiment, the step of determining the violation of at least one other constraint comprises determining the quality of service parameters such as maximum allowable delay or long-term data rate.

According to an advantageous embodiment, the released allocation units are used in rescheduling the available allocation units to another user or service for a particular frame. In this manner, the resources are used in an optimal manner.

According to a preferred embodiment, the result of the scheduling algorithm is signaled to the MT in order to enable it to adapt itself to a possible release of allocation units and to take appropriate measures, for instance, shutting down at least part of its receiver circuitry in order to save power. Preferably, the signaling is transmitted on an associated control channel or broadcast channel in order not to congest the logical data channel.

According to a further embodiment, the steps of checking and releasing allocation units for all users are carried out in a time sequential manner rather than simultaneously for all users. This has the advantage that the released allocation units of a particular user can be assigned first to other users before the check on resource

constraints for other users is carried out. In this manner, an optimal distribution and instantaneous use of all released allocation units can be attained.

The invention also provides a scheduler in a mobile communication system, a base station controller and a mobile communication system having means for carrying out the method of performing the scheduling algorithm described above.

The invention will be described with reference to the following preferred embodiments which will be illustrated with reference to the accompanying figures.

Figure 1 illustrates the cellular concept consisting of one base station and six mobile terminals.

Figure 2 shows an example of a time-division frame structure, where two allocation units form one scheduling frame.

Figure 3 shows an example of a time/frequency-division frame structure, where ten allocation units form one scheduling frame.

Figure 4 shows an example of a time/frequency/code-division frame structure, where eighteen allocation units form one scheduling frame.

Figure 5 illustrates a flow chart of the scheduling algorithm according to the invention.

Figure 6 shows an exemplary scheduling of AUs to users by a prior-art scheduling algorithm.

Figure 7 shows exemplary steps for the scheduling algorithm according to the present invention.

Figure 8 shows an example of a time-frequency-division frame structure where forty-five allocation units form a scheduling frame by a prior art scheduling algorithm for four users.

Figure 9 illustrates the scheduling algorithm according to the present invention for the scheduling frame shown in Figure 8.

Figure 10 shows structural details in block diagram form of a base station implementing the scheduling algorithm according to the invention.

Figures 2 to 4 illustrate the concept of scheduling a frame based on an arbitrary number o f a llocation units e ither i n t he t ime d omain (Figure 2) t he t ime-frequency domain (Figure 3) or the time-frequency-code domain (Figure 4).

As mentioned above, the scheduling is performed in a scheduler usually comprised in the base station o r other p arts o f the non-mobile entities of the communication system.

In Figure 5, the scheduling algorithm is illustrated.

In a first step 10, the scheduler schedules allocation units for a user or service in a scheduling frame as exemplary shown in Figures 2 to 4. The scheduling of resources is done with conventional methods under consideration of scheduling parameters such as long-term minimum data rate, delay constraints etc. Examples for the scheduling of resources is shown in Figures 6 and 8.

In step 20, a check is performed whether the allocation units scheduled to a user in a particular scheduling frame suffice to meet a resource constraint, for instance a short-term minimum resource threshold for one or more mobile terminals.

If so, the scheduling is in order and the algorithm may be restarted for the next user or frame. On the other hand, should the result of the check be negative, i.e. the allocation units do not meet with the resource constraint, the algorithm proceeds with step 30.

In the step 30, it is determined whether the allocation units violate at least one other resource constraint, i.e. more important constraints or quality of service parameters which would b e v iolated b y r eleasing t he AUs. E xamples for s uch m ore i mportant constraints would be a long-term parameter such as delay, data rate etc. Hence, preferably, the fulfillment status of this constraint is additionally monitored over a sliding or fixed time window.

If a violation by releasing is determined in step 30, the scheduling algorithm is restarted for the next user or frame otherwise it proceeds with step 40.

In step 40, the scheduler may release the allocated resources for that particular frame for a user or service provided that this does not violate other constraints or quality of service parameters checked in step 30.

Finally, in step 50, the scheduler reschedules the release allocation units to other user(s) for that particular frame.

It should be understood that the above illustrated scheduling algorithm shown in Figure 5 serves only as an example. In fact, any arbitrary combination of short-term and long-term resource constraints can be made subject to the decision whether an allocation unit for a particular user or service should be released for a scheduling frame. The checking and determination steps on whether a resource constraint is fulfilled or not may be done with a positive or negative result, i.e. whether the resource constraint is met or not. For instance, in step 30, the determination whether another resource constraint is violated could be replaced by a check whether another resource constraint, for example a long-term resource constraint is met additionally and if so allocation units are released in step 40.

A person skilled in the art is able to define a suitable set of resource constraints which need to be fulfilled or violated dependent on the type of user, the system architecture or instantaneous channel or traffic conditions. As an example, in a simple method, step 20 is done for several users simultaneously before resources are released in step 40. A more sophisticated method applies steps 20 to 40 for a first user before these steps are repeated for a second user.

An example for steps 20 to 40 is given in Figures 7 and 9. In the beginning (left part), all AUs are allocated to 4 users in the cell. Applying now step 20, the scheduler finds that the AUs allocated to user 4 are insufficient to meet the minimum resource constraint. Following the rule given in step 40, those AUs previously allocated to user 4 are released again for rescheduling (middle part). Rescheduling assigns the now free AUs to any of the other users, which results in an allocation of AUs to users 1-3 according to the right part of the figure.

It should be mentioned that in this description it is mostly referred to scheduling users. However it is possible to apply the same ideas in a service-oriented (rather than user-oriented) scheduling scenario.

The invention also encompasses the case that not all of the available allocation units are subject to the proposed scheduling approach. Some of the AUs (or users) may be exempt from the checks described in steps 20 and 30 above.

While it has been described above that the scheduler, implemented in the base station determines the resource constraint, the invention also covers cases wherein the resource constraint is given by the communication system design, or is signalled from the MT. In either case, during an ongoing transmission or connection the resource constraint may change its value and a rescheduling may be initiated any time.

Preferably, the minimum resource constraint check is applied to allocation units (e.g. in the time / frequency / code domain) having a quantity of transferable information bits. However, for implementation purposes other quantities may be easier to obtain, calculate or estimate, or may be more representative than bit figures. Such quantities include, but are not limited to a number of modulation symbols, FEC code blocks or Internet protocol packets.

Usually, it is preferred to schedule a large number of adjacent allocation units to a single user. This illustrated in Figure 8 as compared with Figure 6. Typically, a complete frequency band is allocated to a user in order to reduce the amount of signaling which needs to be sent from the base station to the mobile terminal.

As mentioned before, the result of the scheduling algorithm obtained in steps 20 and 30 of Figure 5 may be signaled, preferably in an associated control channel, to the mobile terminal. Hence, the terminal can adjust itself to this situation that it will not receive scheduled allocation units in the current scheduling frame. In response thereto, for example, the mobile terminal can shut down parts of its receiving circuitry in order to save energy.

Figure 10 shows exemplarily some structural details in block diagram form of a scheduler implemented in a transmitter (e.g. either the base station or mobile station) for carrying out the scheduling algorithm according to the invention.

As apparent from the figure, transmitter 1 includes a scheduler 2, a control unit 3 and a check and release unit 4, as well as buffer memory 5 and transceiver circuitry 6. All

other conventional structural details of the transmitter have been omitted as far as they have no direct impact of the invention. The scheduler 2 together with the control unit 3 implements the scheduling algorithm 10 and potentially 50 shown in Figure 5. The checks on the fulfillment or violation of the resource constraint 30 and subsequent potential release of allocation units 40 are performed in the check and release unit 4. Resource constraint values are preferably stored in a memory 5 which is accessed by the control unit 3 to be available to scheduler 2 and check and release unit 4. Alternatively scheduler 2 and check and release unit 4 may have direct access to memory 5. The resource constraint values may be updated upon system initialization, connection setup, or upon receiving a corresponding command from the network controller of the system.

Finally, the transmitter comprises transmission and receiving circuitry 6 for sending and receiving data and control signals using its antenna over the air interface. As mentioned before, preferably signaling data concerning the release of allocation units are transmitted to or received from other transmitting and receiving units of the system. Again, the details of the transmission and receiving operation using logical data channels and control channels is known for the person skilled in the art in the field of communications.

Although the above description has been focused on having the scheduler implemented in the base station acting as transmitter, the principles of the present invention may be readily applied by those skilled in the art to a mobile terminal acting as transmitter for sending data to a base station as the receiving unit, i.e. uplink. In this case, the scheduler is implemented in the mobile station to perform the scheduling algorithm as described above. The resource constraint is either defined by the mobile itself or upon prior instructions from the base station or network controller of the communication system.